TECHNIUM
SOCIAL SCIENCES JOURNAL

9RO 24 1

$ QHZ GHFD
IRU VRFLDO F

,661

ZZZ WHFKQLXPVFLHQFH FRP

Technium Social Sciences Journal
Vol. 24, 113-124, October, 2021
ISSN: 2668-7798
www.techniumscience.com

Reviewing the logic of social scientific claims

Stephen Gorard[1], Yiyi Tan[2]

[1 2]Durham University Evidence Centre for Education

s.a.c.gorard@durham.ac.uk

Abstract. This paper considers three different claims to knowledge, namely, “fully descriptive”, “generally descriptive” and causal claims. These are all common in social science, and each type of claim requires more assumptions than the previous one. After discussing their methodological and logical foundations, this paper describes some of the limitations in the nature of these three claims. Fully descriptive claims suffer from non-random errors and inaccuracies in observations, and can be queried in terms of utility. Generally, in addition to observational errors, descriptive can be questioned because of the long-standing problem of induction. Even the notion of falsification might not be able to help with this. Finally, causal claims are the most problematic of the three. While widely assumed, causation cannot be observed directly. The paper combines and develops three models of what causation might be, and discusses their implications for causal claims. It points out that so far our belief in causation is still a kind of religious one, and that neither theory nor inferential statistics can help in proving or observing its existence. Finally, the paper provides some suggestions for avoiding being misled by false knowledge and reporting our research findings with tentative care and judgement.

Keywords. Causation, Induction, Falsification, Theory, Statistical testing

Introduction

This brief paper concerns the kinds of knowledge that research can generate in social science (and in other fields as well, we assume). It outlines three different claims to knowledge – description, generalisation, and causation – and then considers their methodological and logical foundations. The paper ends by discussing some of the possible implications for causal modelling, the role of theories and inferential statistics, and how it may be possible to help avoid being misled by false knowledge.

For the basis of this discussion we want to consider three different kinds of knowledge used in social science research (and presumably more widely). We will call these “fully descriptive”, “generally descriptive”, and causal claims.

Three kinds of claims to knowledge

Fully descriptive claims are those based on a limited set of observations or data points, where the claims concern only those data points. A simple example might be reporting interview data from a number of participants.

More generally descriptive claims are those based on a limited set of observations, just like fully descriptive claims. But here the observations are used to make a more general claim to knowledge. A common example might be the attempt to generalise the findings from a sample of participants to a wider population of cases not participating in the research.

Causal claims must, we think, also be general claims, based on a limited set of observations. But they add a further conceptual element, by claiming that some observations were created (or modified, influenced or impacted) by other observations. An example might be the proposition that gaining a particular educational qualification leads to a higher-paid job.

Each type of claim requires more than the previous one, mostly in terms of the assumptions made, and the claims are therefore increasingly hard to justify. However, they all have several aspects in common. All start with some observations (or data points of any kind). These observations are the “facts” that underpin each claim. But in social science these would not actually be factual. The observations could be biased, mistaken, misunderstood, misrecorded or misreported.

One can try to minimise these problems with the data, and there is a wide literature on how to do so. It might help if the observations were: automated, replicated, made by people who were unaware of the purpose of the research (blinded), made by people with no vested interest in the results, checked for the “reliability” of several observers, collected about the same phenomena in different ways, and so on. Failure to consider such assistance suggests that the observer is really only interested in what they *think* is going on, and not concerned that others are persuaded by the “truth” of their account.

Nevertheless, whatever is tried, in any reasonably large set of observations there will be errors. A key point to note here is that people have no reason to assume that these errors will be “random” in nature (randomness is discussed further below). Bias, by definition, is not a chance occurrence. Research has long suggested that misrecording or misreporting data is not random, and tends to favour the prior beliefs of the researcher (Adair 1973). This means that researchers have no easy, or even technically complex way, of estimating the scale and impact of such errors, let alone of correcting them. Care and judgement are needed, but something like inferential statistics cannot help. There is no randomisation or probability to assess.

Even fully descriptive claims usually involve more than the observations themselves. There will be some kind of analysis as well. To continue the example above, maybe the researcher will report how many interview participants responded in a particular way, or whether participants with a specific characteristic responded more frequently in a particular way. In the former case, the count of participants itself may be in error. Any form of analysis can be conducted wrongly in practice, or misapplied to the context. And the more complex it is the more likely it is to be in error. In the latter case, the observation of the “specific characteristic” can be in error, just like the observation of the participant response. And just like errors in the original observations, any subsequent analytical errors will not be random in nature. A mistake in counting or in classification cannot be addressed or even identified by any process predicated on randomisation. These problems, and many more, will arise for any empirical claim. Care, simplicity (parsimony, see below), and transparent judgement are the main ways we can think of to deal with them.

In addition, generally descriptive claims require something like induction. These claims have all of the problems with fully descriptive claims, but they are also problematic in making statements about observations that have not been made (and that might never be made). Imagine being handed a large bag containing very many marbles (analogous to a population of cases). It may not be clear exactly how many marbles are in the bag. It may not be clear what colour the marbles in the bag are, or how many there are of each colour. Pulling out several marbles and discovering that they are all blue, for example, does not mean that the others in the bag are all blue. The “several marbles” are analogous to the idea of a sample used in research (Gorard 2021).

While describing a sample in fully descriptive work has some technical problems, it is a relatively simple process in terms of logic. However, going beyond the sample to make assumptions about cases not in the sample has a much shakier logical foundation. It is not clear that revealing several blue marbles from the bag tells us much, if anything, about the remaining marbles in the bag. Imagine pulling the first marble out and discovering that it is blue. What would that tell us about what was in the bag? Precisely nothing. The same for the second and third marbles.

Again, inferential statistics cannot help. If one already knew how many balls there were of each colour in the bag, then one could easily compute the probability of obtaining any combination of colours when sampling from the bag. However, without already knowing what is in the bag, revealing a subset of marbles does not permit one even to say what the probability was of those marbles being revealed – let alone compute the probability of the next marble from the bag being revealed as blue (or any other colour). The sample of marbles does not reveal what colour the other marbles in the bag are. It says nothing about the remaining contents of the bag.

To apply the analogy to research, if we already know what is in the bag then we do not need to do the research to find out. And if we do not know what is in the bag then we can only find out by taking out every single marble. It may seem therefore, that generally descriptive claims can be made about populations, where the entire population is observed (or otherwise has data collected about it). In a sense, this is true. But we prefer to envisage this as being an extreme example of fully descriptive work. There is no generalisation beyond the case actually observed. In any case, social science population data will nearly always be incomplete due to non-response, dropout, or simply missing values about some existing cases. And the problems of miscounting etc. still apply to population data.

Moving to causal claims add a further problematic issue. Now a researcher may want to make a general claim, with all of its attendant problems, but also to state that one aspect of their observations is *caused* by another aspect. This causation is not directly observable. It is not even, like the unobserved cases in a general claim, something that could ever be observed. Instead, it is concept that we use to try and explain regularity in findings. Again, it should not really need to be stated that one cannot demonstrate or prove the existence of a causal model underlying observations through any technical or probabilistic means, such as inferential statistics.

In the knowledge “pipeline”, therefore, from setting out to collect data to positing a causal claim, there are problems of methods and of logic at each stage. And at no stage is traditional probability of any use. The kinds of uncertainty that appear in social scientific claims are not probabilistic at all.

We now consider each of the kinds of claims is slightly more detail.

The relative futility of fully descriptive claims

Perhaps the major problem with fully descriptive claims is why one would want to make them. Research is not journalism or novel-writing. Simply describing the characteristics or experiences of a limited number of cases is not often useful. Readers would immediately want to know if these findings are special or permanent or true more widely. They would want to know the usefulness of the findings. Of course, a larger number of such studies could be conducted, and then combined to create a larger dataset. This does not always help, but it can do. For example, the term statistics derives from knowledge about the state. Political arithmetic was/is the simple descriptive portrayal of "facts" about the state, such as levels of poverty, ill health, and infant mortality. This approach can lay bare a problem in a way that is hard for politicians and others to ignore. It can be invaluable. Nevertheless, it is just a start. Even with political arithmetic, readers are quickly moved to ask whether these figures are equally true everywhere, why they arise, and what should be done to ameliorate them. These more complex but interesting questions cannot be answered by mere description.

The problem of induction in generally descriptive claims

If one wants to make a more general claim to knowledge – taking a Popperian example (see below) – that all swans are white, then observing one white swan is not enough to sustain the claim. Even 100 or one million observations is not enough. Replicating observations in this way does not seem to help establish the general claim. All one can truly do, even after observing one million white swans, is to make the fully descriptive claim that one million swans are white. And, as above, even this claim is not absolutely clear, because of the possibility of misclassifying, misrecording and so on. Any more general claim would have to be tentative unless or until one has seen and judged the colour of *all* swans. And how would one know, in practice, that they had seen all swans?

In the same way as Popper's example, all of our claims based on research data are limited (even where they have been replicated and peer-reviewed). They must be seen as very tentative. With induction, the replication of our results is not actually that important. One can see many white swans without the claim being true, and a research "finding" can be replicated many times and still be wrong.

Hume (1962 edition) introduced this "skeleton in the cupboard of philosophy" - that the process of inductive reasoning has no logical foundation. Yet induction has often been used as the chief criterion of demarcation between what is considered "science" and what is not.

Popper (2002 edition) suggested a way around this, by highlighting the notion of falsification. This kind of testability, he said, was the true difference between science and all else. One cannot, for example, conclude with logical certainty that all swans are white merely from repeated observation of white swans (induction). But one can falsify the claim that all swans are white by just one observation of a non-white swan. Thus, progress comes from falsifying theories not from further confirmation of them. This is an attractive idea, and it chimes with the writing of Mill in relation to identifying causal models (see below). But is it true that Popper's falsification evades the use of induction?

In formal logic, the statements "A entails B" and "Here is an A which is not B" form a contradiction. Neither can be said to falsify the other because one would not know which, if any, of the statements is true. One only knows that both cannot be true. There is no logical justification for saying that the example of "A which is not B" means that "A entails B" is false. Since A and B are ideal terms we do not attempt to tinker with them and overcome the contradiction. Contradiction is *not* the same as falsification.

The idea of falsification arises from the fact that these nouns and adjectives are not logical entities. They are names for real-world things, and in that real world there is bias, misclassifying and so on. In the real world, where A and B become swan and white, one can at least consider the possibility that only one of the propositions is falsified by the contradiction. This is what Popper does without making this step explicit. He then states that it is clear which proposition is wrong - so clear that the alternative is usually dismissed as merely playing with words (Thouless 1974). But this supposed clarity is, like induction, actually only a habit of mind (see below).

In the example, Popper proposes that we change the definition of swan to include the possibility that some swans are not white, and does not even bother to argue against the alternative. Nevertheless, the other way out of the contradiction is equally *logical*. We could change the definition of black to exclude the possibility of being applied to swans. Thus, the thing that looks like a black swan is actually not a swan because it is black. The choice is between changing our definition of swan or of black. In this example, people prefer changing the definition of the least familiar term, and black is a much more familiar term than swan. If the same is true in every example of falsification then what seems like a logical argument for falsification is actually an appeal to the same non-logical phenomenon of familiarity that underlies induction. When observation leads us to question a belief because it brings two beliefs into contradiction people tend to stick with the most familiar of the two concepts. This suggests that Popper's notion of falsification does not actually eliminate inductive logic at all (see also Goodman 1973). Familiarity breeds certainty in a way that is logically unjustified.

Take another example of a claim – that all doors are rectangular in shape. Many doors are rectangular, partly because people can control their shape. But some are not – like in igloos, or the International Space Station. There will also be cases of different shapes that we are not sure are doors. The problem, as with the "bias" in falsification, lies in the use of words. Words are not like the categorical algebraic or logical variables A, B etc. Our name categories, like door or swan, are imposed on things that could really be continuously variable in nature. One might, in theory, line up everything in the world, in order of "doorness", and somewhere there would be examples that are genuinely hard to classify (maybe some kinds of windows). We all know this. As Russell (1903) and others have shown, putting things into clearly delineated sets may not work. So, the induction problem is not one of knowledge *per se* but stems mostly from language and the use of categories.

The habit of causation

Causal claims are even more problematic than general descriptions. Causes and effects are ideas used to describe a firm impression people have about the way the world works. Events and processes have a regularity and time sequence that offer both an explanation for why things occur, and even a way of controlling them. Social science, perhaps more than other fields, is pervaded by what Abbot (1998, p.149) called 'an unthinking causalism", which appears to be worsening over time (Robinson et al. 2007). Correlations, patterns and even just perceptions and opinions are routinely presented by researchers in very definitive causal terms. We all need to be clearer about what it means to make the strong claim that something causes something else.

Hume (1962) described cause and effect as an immutable habit of mind – people are pre-disposed to see regularities in their environment and ascribe something like causation to them. This may have been a valuable evolutionary heuristic when time was short and a quick decision was needed. But it can lead to mistakes and superstition in the longer term. Across his different writings Hume seemed to be somewhat ambivalent about causation (Coventry 2008). On the one hand, as a "matter of fact", all that one has to support the existence of causation is the observed regularities of nature. One cannot use Hume's "relation of ideas" to deduce causation logically from any such available facts or regularities. But Hume also suggested that causal claims are, and must be, testable propositions about knowledge.

Causes cannot be deduced just from observing effects (Blalock 1964). Seeing a light bulb going off does not, by itself, allow the observer to deduce whether it has been switched off, or there is a power failure, or the bulb is broken, for example (Salmon 1998). Similarly, effects can rarely, if ever, be deduced simply by observing their possible causes. Who would

Technium Social Sciences Journal
Vol. 24, 113-124, October, 2021
ISSN: 2668-7798
www.techniumscience.com

have thought before experiencing it that striking a flint could create a fire, for example? How could we tell what an unknown switch might turn on?

Potentially, causal models are very complex. Any event could be the effect of a large number of contributing causes. All of these causes might be needed to create the effect, but be insufficient in isolation. All causes might work only within a given context, or only in combination (Emmet 1984). A fire needs oxygen, flammable material, and ignition (a flame). One can say that the flame causes the fire, but it does not do so alone, and a variety of causes could be sufficient to create the effect, with none of them strictly necessary. One might start a fire with a lighter, a match, a flint, or a magnifying glass. Also, any cause or combination of causes could have more than one effect. Starting a fire causes combustion of the flammable material, but it also causes heat and light.

These issues are all problematic for Hume's idea of cause and effect having constant conjunction. If C is caused by both A and B in combination, then the correlation between A and C in isolation may be zero. The same thing arises for B and C in isolation. We may therefore be unable to predict exactly what the effects of a set of causes might be, because of the complexity of their interaction. Instead, one might predict their effects in probabilistic terms, or after controlling for everything else. An example of a *ceteris paribus* causal model could be the erosion of a river bank caused by a meandering river (Corbi and Prades 2000). There is no doubt that the river bank will erode over time even though it is not possible to be precise about the exact pattern it will form. This is reasonable, but makes it hard to test any causal model in practice.

There have been many attempts, since Hume, to describe the elements needed to establish a strong causal claim. For Mill (1882) a cause has three key elements. It should be clearly related to the effect (correlated through observation), it must precede the effect, and there must be no plausible alternative explanations for the effect other than the cause.

The first of these elements, the association of the putative cause and its effect, is certainly a *sine qua non*. Commentators might say that a correlation is not the same as causation, but not having a correlation between two things surely means that neither is the cause of the other. So one can test a causal claim to the extent that one can assess a correlation as part of a generally descriptive claim (see above).

Of course, assessing such a correlation may not be easy in practice. In some of the natural sciences one might clone cells, or find identical particles. Hume considered billiard balls, which are also similar to each other, and may be envisaged as interchangeable. In social science, however, one cannot usually expose the same people or organisations both to a research process and not. This means that the results of causal research in social science is not generally clear-cut. People might use statistical approaches to express the nature of causal models, and this may lead others to imagine them as being probabilistic (Goldthorpe 2001). But actually, they reflect the limitation of our understanding, not the reality of the world (Shafer 1996).

Viewing causation as a stable association between two phenomena, as Mills and Hume do, creates several problems. It is clearly wrong to suggest that a singular event cannot have a cause or causes, but there can be no repeated association between singular events – such as the onset of the Second World War. In a sense, all events are singular in terms of time, place, context, and actors involved. Mills' criteria are best understood as describing how one can identify causes, and are not necessarily characteristics of all cause:effect sequences. Where one can observe or repeat very similar situations, such as striking a billiard ball in Hume's account, it is much easier to test a proposed causal model than when faced with a complex causal question about a one-off process, such as what caused the outbreak of the Second World War.

In both classical and operant conditioning, it has been shown that the association of two things leads the conditioned subject to behave in the presence of one thing as though it implied the presence of the other (Skinner 1971). Skinner's pigeons "learnt" to pull a lever which had always accompanied the release of a pellet of food in the past. The conditioned subjects do this whether the lever is mechanically releasing the pellet or not. To an observer, the pigeons behave as though the lever is a cause. In intermittent reinforcement schedules, where the pellet appears on only some occasions, this behaviour is even stronger – it will take more examples of no pellet after pulling the lever to un-condition the subject than it would if the pellet had previously always appeared.

Further, Skinner's accidental reinforcement schedule is a powerful reminder of the dangers of allowing causal models to be based only on association. In accidental reinforcement schedules, providing pellets at random tends to reinforce whatever behaviour the subject was involved in at the time. That behaviour is then more likely to be repeated by the subject, and so more likely to coincide with the next random arrival of a pellet. The more the pigeons perform the ritual the more likely it is that food will randomly follow one of the performances. This continuously reinforces the ritual. Eventually, the subject repeats an endless superstitious ritual of one behaviour, only intermittently reinforced by the arrival of a pellet, so making the apparent association resistant to un-conditioning. The response becomes self-fulfilling.

These findings suggest that the kind of imagined probabilistic causation, commonly reported in social science, will paradoxically be an even stronger habit of mind than Hume's constant conjunction idea. And this is so, even though it is actually more likely to be an erroneous association than a constant conjunction would be, partly because of the complexity of deciding whether a purported cause that is only sometimes "effective" is actually a cause at all. And partly because it may be accidental (a superstition). Our task as researchers is to avoid such superstitions as far as possible.

Mill's second element is also problematic. It is not necessarily true that a cause must precede an effect. They can be contemporaneous. Some observations which are seemingly in a temporal sequence may actually be reciprocal (Hagenaars 1990). One can accept causes simultaneous with their effects, such as where a ball rests on a cushion, and the cushion is causing the ball not to drop further (Mackie 1974). If we drop two balls into a bowl, we can model the final resting places of both balls mathematically, but we cannot use this to decide which ball is "causing" the other to be displaced from the centre of the bowl. The events are surely mutually determined (Garrison 1993). Mathematical statements or systems of equations can describe such systems but they cannot express either intention or causality. They can be used to show that systems are, or are not, in equilibrium, and to predict the actual change in the value of one variable if another variable is changed. However, it is important to recall that this prediction works both ways. If $y=f(x)$ then there will be a complementary function such that $x=f'(y)$. Which variable is the dependent one (on the left-hand, predicted side) is purely arbitrary. Nothing in mathematics, logic, or statistical analysis can overcome this limitation.

In fact, all one can say, with some conviction, is that our present models do not permit a reverse sequence of causation. The effect cannot come before its ultimate cause. Student attainment at age 16 cannot cause their attainment at age 11, in any real sense.

Mill's third element is the need for an explanation. It is correct that such an explanation must be the simplest and most plausible. A causal explanation describes a process that shows how the cause could generate the effect. A good explanation must be easy to test, and must make the fewest assumptions necessary to provide a mechanism linking cause and effect. The proposed effect must be capable of change, and it must be capable of being changed by the proposed cause (de Vaus 2001).

A good example is the clear relationship between smoking and lung cancer. The statistical conjunctions and the observations from laboratory trials with animals were explained by the isolation of carcinogens in the smoke, and the pathological evidence from diseased lungs. These combined to create an explanatory theory.

However, it is not clear that an explanation is essential to a causal claim. It is possible to switch a light on and off without understanding how it works. The fact that it does work is part of what shows that the switch is the cause of the light going on and off. This suggests that the explanatory mechanism is the least important part of any causal model. If it is clear that altering the presence or strength of a potential cause works in the sense of changing an effect, then it matters less if the mechanism is fully understood or not. And, of course, even the most convincing explanation possible is of little consequence if the potential cause has no discernible effect in practice.

Bradford-Hill (1966), and others working on the links between smoking and lung cancer, proposed a somewhat tougher set of scientific conditions for the identification of a causal link than Mill. Some of these are clarifications of Mill's elements, establishing rules for how and when Mill's elements will have been established. For example, the first element, correlation, must be found in different studies, led by different researchers, using different methods and differing cases. This additional specification is good practice, but is not a philosophical component of a causal model. In addition Bradford Hill tried to address the lack of constant association in some contexts (see above), by saying that the frequency of association between the cause and effect must be substantial compared to the frequency of either X or Y in isolation. They no longer have to be constantly conjoined. This is appropriate but it does make the identification of causes harder.

Mill's sequence element is divided into two by Bradford-Hill. So, a cause must be able to predict the effect (as discussed above). And the cause must come before the effect (but see limitations of this idea for contemporaneous events, above).

The third element is again a coherent, plausible, workable mechanism explaining how the cause can influence the effect. But it should also be "agreed" and "consistent with prior knowledge". Again, these additional requirements sound sensible for practice (or in a legal case), but again they do not form part of the logic of causes. Something could be correctly identified as a cause by only one person. And that identification might create a scientific revolution that is not consistent with what was assumed to be prior "knowledge". Also missing from the Bradford-Hill account is the "elimination of sensible alternative explanations". This elimination can be through robust testing of all possible explanations, or it can be based on an argument such as that the best explanation is the most parsimonious one. Plausibility is not enough for a theory.

Bradford-Hill adds a fourth element in two parts. There must be a reduction in the effect after removing or reducing the cause. And there must be an increase in the cause after the introduction of, or increase in, the cause. This is useful. It adds a requirement that deliberate variation in the appearance or strength of the cause must yield a change in the effect. Put another way, one must not consider a causal model established unless it has been robustly tested (though an experimental design or similar). However, it still assumes the idea of a constant conjunction, and that the effect has only one cause. Neither is necessarily true.

A key point is not whether one can explain why a cause has an effect, but whether it can be demonstrated to have an effect at all. Causation can best be viewed via the impact of an intervention. Does the proposed causal model work in practice, under controlled and rigorously evaluated conditions? Since causes are not susceptible to direct observation, but what they cause is effects, then at least those effects must be observable (like a light coming on, when the switch is pressed). We need evidence that controlled interventions have altered the level or presence of the potential cause, and so produced changes in the purported effect that cannot be explained in any other way.

Gorard (2002, 2013) reformulated these elements into a simplified model of causal evidence for social science, consisting of four main criteria. These criteria do not require *constant* conjunction. They allow cause and effect to appear at the same time (but not with the cause after the effect). They include the need for intervention studies. And they insist that the explanation must be warranted by the full body of evidence available.

1. For X (a possible cause) and Y (a possible effect) to be in a causal relationship they must be repeatedly associated. This association must be strong and clearly observable. It must be replicable, and it must be specific to X and Y.

2. For X and Y to be in a causal relationship, they must proceed in a suitable sequence. X must always precede or appear with Y (where both appear), and the appearance of Y must be safely predictable from the appearance of X.

3. It must have been demonstrated repeatedly that an intervention to change the strength or appearance of X clearly changes the strength or appearance of Y.

4. It helps to have a coherent mechanism to explain the causal link. This mechanism must be the simplest available without which the evidence cannot be explained. Put another way, if the proposed mechanism were not true then there must be no simpler or equally simple way of explaining the evidence for it.

Discussion

In this paper, we have been talking about claims to knowledge. It is perfectly proper for researchers to speculate about their results, and possible meanings. But it must be clear to their readers that these are speculations and not claims to knowledge. There is a key difference.

Should we be concerned about causation?

There can be no direct evidence that observations are either caused, or somehow just random events that might seem patterned, in the same way as there are sequences in a table of random numbers (Arjas 2001). Either explanation fits the facts. So using either as an explanation for observed phenomena involves making an assumption not contained in any data. To use both to explain observations involves making *two* assumptions, and is therefore un-parsimonious. It is hard enough to establish whether causes exist or not. To allow them to exist alongside unrelated phenomena would make most social scientific propositions un-testable (Gorard 2013).

Given a choice, it is consistent for anyone conducting social science research to assume that causation exists, else why would they bother to the research? It is the same argument that could be made about the enervating implications of solipsism. Researchers are logically and ethically required to accept that causation is a real possibility. If we genuinely reject the idea of causation then research, and trying to improve social conditions, become pointless. The latter can only make sense if researchers can make a difference (i.e. have an effect). However, our preference for causation (invisible and without form) is still no more than a kind of religious belief.

Implications for theory and inferential statistics

It is clear that theory or explanation is the least important element of a causal claim (or indeed any claim). If two items are unrelated to each other then neither can be the cause of the

Technium Social Sciences Journal
Vol. 24, 113-124, October, 2021
ISSN: 2668-7798
www.techniumscience.com

other. If the apparent effect appears before the apparent cause then the causal claim is considered wrong. If varying the cause never produces a change in the apparent effect then the causal claim is wrong. But the explanation is not about what is true. It is about people's understanding of what is true. Just as a cause can exist without anyone noticing it, so a cause can exist even if no one understands how it works. The fact that it works (or at least appears to work) is enough. Theory is often playing catch up, in trying to explain new findings, as well as sometimes generating new ideas to test.

The simplest explanation of any observation(s) is the best not because it is proven or more likely, but because it is easier to test than any more complex ones. Adding needless elements to the explanation makes it confusing. Explanation must be trimmed down to the minimum needed, and this also makes them easier to test. Maybe this is what is happening with the preference for changing the definition of swan rather than black, in the verbal form of a syllogism. Maybe it seems simpler to test the colour of swans than to test what things colours should or should not be attached to. This needs for simplicity leads to a warranting principle (Gorard 2002). Before drawing a conclusion, see whether the observations (data/evidence) can be explained at least as well by any simpler conclusion.

If the above argument is accepted then it is also an illusion that inferential statistics can be used to help make any social scientific claims. None of the problems identified in each kind of knowledge claim is probabilistic in nature. From the early discussion about errors in collecting and assembling data to the lack of a logical foundation for causal claims, the kind of uncertainty introduced in all knowledge claims cannot be addressed by significance tests (or confidence intervals, and related paraphernalia).

Care in making/establishing claims

Given the rather unsatisfactory nature of all knowledge claims, we suggest that the best way to help avoid being misled lies in care and good judgement. All claims are clearly contingent. A claim that is unfalsifiable useless and may be damaging.

Even logical claims are suspect when the elements in an argument refer to real-life things, and are not solely symbolic. Ignoring the implications of Gödel's incompleteness demonstrations (Smullyan 1992), logical deductions are generally considered to be necessarily true (and 2+2 must equal 4, by definition). If we assume that all swans are black then it must be true that none are white. This would be logical. And must be necessarily true within the context of the assumption. But these are tautologies, and they contain no information of the kind that would reduce uncertainty as defined in Information Theory by Shannon and Weaver (1949). However, in real-life research we may miscount to five, or misclassify a swan etc. Since our data is always tentative, we can never be sure of any claim or finding, even though the analysis might be perfect.

Transparency of data and analysis help. They help others understand how the data was assembled and its flaws, and to judge whether to trust that the research is the best bet for them to act on for the present.

A fully descriptive claim must be based on the most accurate observations possible, even where there is no intention of moving to a more general claim. For example, if a claim is made that one group of cases with data differs from another group, then this claim must be internally valid (i.e. it must be "true" of the cases in the study).

There is no point at all in trying to establish whether an invalid or untrue claim is then more generally true. It has not even been established as a valid fully descriptive claim. Generalisation is something that is only relevant once a secure claim has been established. As discussed, there is no way of knowing for certain whether a secure finding is also true of other cases not involved in the research (whether, indeed, *all* swans are white). Attempted falsification, in imaginative ways, might help but as we have illustrated there are problems even in that. And, as above, statistical analysis must be silent on this.

Moving from merely descriptive claims to causal ones introduces more barriers to security. This means that the move in research from one to the other should start with the securest claims and ensure that a similar regard to the safety of findings is applied to every stage.

Maybe the problem with general propositions, such as that all swans are white, is that we are looking at them wrongly. Or maybe they are not the kind of propositions that we should be seeking to establish anyway. Perhaps another problem is the descriptive or reactive nature of general propositions used as examples. Maybe the focus should be more on statements about *action* – of the kind if we do X then we expect Y result.

References

[1] Abbot, A. (1998) The causal devolution, *Sociological Methods and Research*, 27, 2, 148-181
[2] Arjas, E. (2001) Causal analysis and statistics: a social sciences perspective, *European Sociological Review*, 17, 1, 59-64
[3] Blalock, H. (1964) *Causal inferences in nonexperimental research*, Chapel Hill: University of North Carolina Press
[4] Bradford-Hill, A. (1966) The environment and disease: Association or causation?, *Proceedings of the Royal Society of Medicine*, 58, 285
[5] Corbi, J. and Prades, J. (2000) *Minds, causes, and mechanisms*, Oxford: Blackwell
[6] Coventry, A. (2008) *Hume's theory of causation*, London: Continuum
[7] de Vaus, D. (2001) *Research design in social research*, London: Sage
[8] Emmet, D. (1984) *The effectiveness of causes*, London:Macmillan Press
[9] Garrison, R. (1993) Mises and his methods, pp.102-117 in Herbener, J. (Ed.) *The meaning of Ludwig von Mises: contributions in economics, sociology, epistemology, and political philosophy*, Boston: Kluwer Academic Publishers
[10] Goldthorpe, J. (2001) Causation, statistics, and sociology, *European Sociological Review*, 17, 1, 1-20
[11] Goodman, N. (1973) *Fact, fiction and forecast*, New York: Bobs-Merrill
[12] Gorard, S. (2002) Fostering scepticism: the importance of warranting claims, *Evaluation and Research in Education*, 16, 3, 136-149
[13] Gorard, S. (2013) *Research Design: Robust approaches for the social sciences*, London: SAGE
[14] Gorard, S. (2021) *How to make sense of statistics: Everything you need to know about using numbers in social science*, London: SAGE
[15] Hagenaars, J. (1990) *Categorical Longitudinal Data: Log-linear, panel, trend and cohort analysis*, London: Sage
[16] Hume, D. (1962) *On human nature and the understanding*, MacMillan
[17] Mackie, J. (1974) *The cement of the universe*, Oxford: Clarendon Press
[18] Mill, J. (1882) *A System Of Logic, Ratiocinative And Inductive, being a connected view of the principles of evidence, and the methods of scientific investigation* (8th Edition), New York: Harper & Brothers
[19] Popper, K. (2002) *The logic of scientific discovery*, Routledge

[20] Robinson, D., Levin, J., Thomas, G., Pituch, K. and Vaughn, S. (2007) The incidence of 'causal' statements in teaching-and-learning research journals, *American Educational Research Journal*, 44, 2, 400-413
[21] Russell, B. (1903) The principles of mathematics, New York: W. W. Norton & Company
[22] Salmon, W. (1998) *Causality and explanation*, New York: Oxford University Press
[23] Shafer, G. (1996) *The art of causal conjecture*, London: MIT Press
[24] Shannon, C. and Weaver, W. (1949) The mathematical theory of communication, University of Illinois Press
[25] Skinner, B. (1971) *Beyond freedom and dignity*, Hackett Classics
[26] Smullyan, R, (1992) Gödel's Incompleteness Theorems, Oxford: Oxford University Press
[27] Thouless, R. (1974) *Straight and crooked thinking*, London: Pan